

10/530, 953

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 August 2001 (23.08.2001)

PCT

(10) International Publication Number  
**WO 01/61684 A1**

(51) International Patent Classification<sup>7</sup>: **G10L 19/00, H04J 1/02**

(74) Agent: NICHOLAS, Frank, C.; Cardinal Law Group, 1603 Orrington Avenue, Suite 2000, Evanston, IL 60201-5043 (US).

(21) International Application Number: **PCT/US00/31338**

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date:  
15 November 2000 (15.11.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:  
09/507,084 18 February 2000 (18.02.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): FIRST INTERNATIONAL DIGITAL, INC. [US/US]; 135 West Central Road, Schaumburg, IL 60195 (US).

Published:

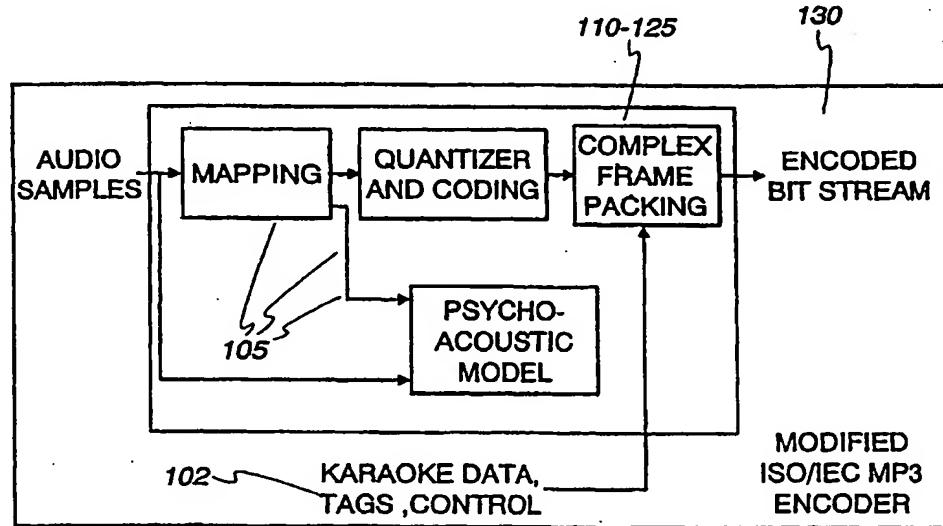
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND SYSTEM FOR ENCODING AN AUDIO SEQUENCE WITH SYNCHRONIZED DATA AND OUTPUTTING THE SAME



BEST AVAILABLE COPY



WO 01/61684 A1

(57) Abstract: A method relates to encoding an audio sequence (100) with synchronized data (102). An audio sample and a data sample are provided. The audio sample is converted into an audio signal. The data sample is converted into a data signal. The data signal includes a plurality of data segments. The audio signal is encoded (110-125) with the data signal to form an audio sequence, including frames with a field for receiving a data segment or receiving pointer signals to a data segment. For output, a compressed data signal, being synchronized with a compressed audio signal in the sequence (200) is unpacked (205-215) and output (220).

5           **METHODS AND SYSTEM FOR ENCODING AN AUDIO SEQUENCE  
WITH SYNCHRONIZED DATA AND OUTPUTTING THE SAME**

**FIELD OF THE INVENTION**

10          The present invention relates to audio sequences, and, more particularly, to the encoding of an audio sequence with synchronized data, and the output of such an encoded file.

**BACKGROUND OF THE INVENTION**

15          Karaoke is a musical performance method in which a person (i.e., the singer) performs a musical number by singing along with a pre-recorded song through the reading of that particular song's lyrics, which are preferably displayed on a display device, such as, for example, a television screen situated within view of the singer. The singer's voice overrides the voice of  
20          the original singer of the pre-recorded song. A video motion picture, often referred to as a music video, may also typically be displayed as an accompaniment to both the music and the singer. Devices providing this opportunity are known as karaoke musical reproduction devices, and will be referred to as karaoke devices.

25          Current karaoke devices use tapes, compact disks (CDs), digital videodisks (DVDs), computer disks, video compact disks (VCDs) or any other type of electronic medium to record and play both the music and the lyrics. With the rise in popularity of karaoke as an entertainment means, more and more songs are put in karaoke format. As a result, the need to transport and  
30          store these ever-growing musical libraries has become paramount. In some instances, digitized data representing the music and the lyrics has been

- 2 -

- compressed using standard digital compression techniques. For example, one popular current digital compression technique employs the standard compression algorithm known as Musical Instrument Digital Interface (MIDI). U.S. Patent No. 5,648,628 discloses a device that combines music and lyrics 5 for the purpose of karaoke. The device in the '628 Patent uses the standard MIDI format with a changeable cartridge which stores the MIDI files.

The International Organization for Standardization (ISO/IEC) has produced a number of generally known compression standards for the coding of motion pictures and associated audio data. This standard is referred to as

- 10 the MPEG (Motion Picture Experts Group) standard. The MPEG standard is defined in documents ISO/ IEC 11172 (which defines the MPEG 1 standard) and ISO/IEC 13818 (which defines the MPEG 2 standard), both of which are incorporated herein by reference. Another popular, non-standard compression algorithm, which is based on the MPEG 1 and MPEG 2  
15 standards, is referred to as MPEG 2.5. These three MPEG versions (MPEG 1, MPEG 2, MPEG 2.5) will be collectively referred to as "MPEG 1/2." U.S. Patent No. 5,856,973 discloses a method for communicating private application data along with audio and video data from a source point to a destination point using the MPEG 2 format.  
20 MPEG 1/2 is further broken into a number of "layers." In general, the higher an MPEG 1/2 layer is labeled, the more complexity is involved. MPEG 1/2 Layer III (MP3) is an emerging popular compression format, which may be used for encoding audio data in an effort to produce near-CD quality results.

- 25 MP3 players are portable devices, typically containing a "flash" memory, a liquid crystal display (LCD) screen, a control panel and an output jack for audio headphones and other similar devices. Musical compositions are loaded into the "flash" memory of the MP3 player through connection to a personal computer (PC) or other similar device, and played for personal enjoyment.

- 3 -

The MP3 standard defines an "audio sequence," which is broken down into variable size "frames," which are further broken down into "fields." Although the syntax of each frame is described in the MP3 standard, the content of the fields within each frame is not defined and is the subject of the  
5 present invention.

Typical karaoke devices are large, complex expensive systems used in bars and nightclubs. They involve large display screens, high fidelity sound systems and a multitude of storage media, such as, for example, CDs.

Typical MP3 players are small and affordable, but are designed to simply play

10 music. They have small display screens to display only the title and play time of a song, limited audio output to a headphone, and minimal (if any) microphone.

Typical MP3 players do not currently possess the ability to synchronize a data field, containing lyrical information of a song, with an audio signal,  
15 containing the musical aspect of the song, into a single audio sequence file that can be stored, manipulated, transported and/or played via a karaoke player device.

Accordingly, it would be desirable to have a program and method that overcomes the above disadvantages.

20

- 3 -

- 4 -

**BRIEF DESCRIPTION OF THE DRAWINGS**

**FIG. 1** is a block diagram illustrating the syntax of the MP3 audio sequence, as described in the MP3 specification standard;

**FIG. 2** is a schematic diagram of an MP3 encoder, as described in the 5 MP3 specification standard;

**FIG. 3** is a schematic diagram, illustrating a modified MP3 encoder, in accordance with the present invention, to embed karaoke data with an audio signal to form an MP3 audio sequence;

---

**FIG. 4** illustrates a flow chart of the encoding process, in accordance 10 with the present invention;

**FIG. 5** is a schematic diagram of an MP3 decoder, as described in the MP3 specification standard;

**FIG. 6** is a schematic diagram, illustrating a modified MP3 decoder, made in accordance with the present invention, to un-embed karaoke data 15 and an audio signal from an MP3 audio sequence;

**FIG. 7** illustrates a flow chart of the decoding process, in accordance with the present invention; and

**FIG. 8** illustrates a block diagram showing the MP3 karaoke player apparatus.

20 Corresponding reference characters indicate corresponding parts throughout the several views. The exemplifications set out herein illustrate one preferred embodiment of the invention, in one form, and such exemplifications are not to be construed as limiting the scope of the invention in any manner.

**DETAILED DESCRIPTION OF THE  
PRESENTLY PREFERRED EMBODIMENTS**

- In the present invention, a preferred embodiment for encoding an audio sequence with synchronized data takes place according to the MP3 standard, as described above. Alternatively, the encoding process described below may be performed according to the confines of other similar standards. These other standards may include, for example, MPEG 1/2 Layer III, AC-3, Microsoft's Windows Media Audio file (WMA), Advanced Audio Coding (AAC), Lucent Technology's Perceptual Audio Coder (EPAC), Liquid Audio, real.com's G2, and other frame based audio format standards. For purposes of this invention, MPEG 1/2 Layer III means MPEG 1, MPEG 2 and MPEG 2.5 Layers 1 and 2 formats. Therefore, the present invention is applicable to any frame-based audio format.
- As mentioned above, the MP3 standard defines an "audio sequence." A typical audio sequence of the MP3 standard is illustrated in FIG. 1. The audio sequence 10 (shown in more detail in of FIG. 1-A) is broken into variable size "frames" 12. An example of one frame of the audio sequence is shown in FIG. 1-B.
- Each frame is then further broken down into a plurality of fields 14 and sub-fields 16. Examples of some of the fields 14 and sub-fields 16 of the frame 12 shown in FIG. 1-B are illustrated FIGS. 1-C, 1-D and 1-E. In the preferred embodiment, each frame 12 of the audio sequence 10 includes a fixed format made up of a header field, an error check field, a main data field and an ancillary data field. Furthermore, each of the fields 14 are broken down further into sub-fields 16, an example of which is shown within the divisions of FIGS. 1-C, D and E. Although the syntax of each frame 12 is described in the MP3 standard, the content of both the fields 14 and the sub-fields 16 within each frame 12 are not defined within the MP3 standard. In addition, the private bits defined in both the header and the audio data frames, as well as the ancillary data frame, can be used to encode lyrical data

- 6 -

and control signals, or cues to lyrical data and control signals, within the audio sequence 10, such that it is synchronized with the audio signal upon the formation of the audio sequence 10.

- It is important to note that the header fields for each frame 12 occur 5 within a fixed period and are a specific size. The data fields associated with each frame 12, however, are of variable size and do not occur within a fixed period.

- More particularly, the present invention concerns using the private bit in the header field (FIG. 1-E, Field 8), the private bits in the main data field 10 (FIG. 1-C, Field 2) and the ancillary data field (FIG. 1-D) to be embedded with lyrical text, video, cues to lyrical text or video, and/or control information. This information will be collectively referred to as karaoke data. It should be noted that each frame may or may not include any karaoke data.

- If a frame does include karaoke data, such data may be stored within 15 any or all portions of the available data fields mentioned above. Preferably the above-described information will be stored within the data fields in the following order: first, the private bit in the header field; second, the private bits in the main data field; and third, the ancillary data field.

- FIG. 2 shows a high level block diagram of an MP3 encoder as 20 described in the MP3 specification. As mentioned above, karaoke data may be encoded in the private bit of the header field, the private bits in the main data field, or within the ancillary data. FIG. 3 illustrates a high level block diagram of a modified MP3 encoder used to encode the karaoke data. The frame packing stage of the encoder must be enhanced to synchronize 25 incoming audio data with karaoke data to pack the frames accordingly. This is done by sending in tags and control information with the karaoke data. The "complex frame packing" unit uses this information to sequence the karaoke data with the audio samples appropriately. FIG. 4 illustrates a flow chart detailing the encoding process of the present invention, with a focus on frame 30 packing the karaoke data. Additionally, FIG. 5 illustrates a high level block

- 6 -

diagram of an MP3 decoder, as described in the MP3 specification. FIG. 6 illustrates a high level block diagram of a modified version of the MP3 decoder. FIG. 7 describes a flow chart of the decoding process with a focus on karaoke data unpacking. During the decoding process, the karaoke data 5 is produced during the frame unpacking stage while the audio data is produced as a final product of the inverse mapping stage. The karaoke data is then sequenced with the audio data external to the decoder.

With reference to FIGS. 1-4, a method of encoding an audio sequence is provided for, as follows. According to the present invention, an encoder

- 10 receives both an audio sample and a data sample (step 100). Preferably, the encoder is a system that is developed to synchronously encode an audio sample with a data signal, creating an audio sequence. In the preferred embodiment, the audio sample is a musical composition. Alternatively, the audio sample may be an oral signal, such as, for example, an audio version 15 of a text, such as, for example, a book, a newspaper or a foreign language textbook. In the preferred embodiment, the data sample may be the words to a musical composition. Alternatively, the data sample may be an oral version of a text, such as, for example, an audio version of an English language text or video data, corresponding to, for example, a music video of the song 20 embodied in the audio sample.

After receiving the audio sample and the data sample, the encoder then converts the audio sample into an audio signal (not shown). Preferably, the conversion process assures that the audio signal will be able to be read and understood according to the preferred format of the audio sequence. For 25 example, if the format of the audio sequence is MP3, then the audio signal will preferably be able to be read according to the MP3 format.

In much the same way, the data sample is converted into a data signal (step 102). Further, the data signal may include a plurality of data segments. Each of the data segments preferably corresponds to a portion of the data sample, such that it may be embedded into the resultant audio sequence.

- 5 Not all portions of the data signal need be encoded within the data segments. Rather, each of the data segments may contain a fractional portion of the data signal corresponding to the data signal.

For example, if the data sample contains the words to a song, the data signal would include various data segments, each segment corresponding to,

- 10 for example, a word or a beat. The purpose for this, which will be described in more detail below, allows the data segment to be embedded into the audio sequence, both in an order and in a location such that the data signal corresponds to the audio signal (i.e., in such a manner that the data signal is synchronized to the audio signal).
- 15 The data signal may also include a control signal. Preferably, the control signal contains information relating to the order of embedding of the data signal within the audio sequence. For example, the control signal may dictate that, during the encoding process, one particular word of the lyrics contained within the data signal may contain three syllables, each syllable
- 20 requiring position at a different beat of the song. Such information would be preferably contained within the control signal.

- After converting both the audio signal and the data signal, the audio sequence is then encoded. The audio sequence consists of the audio signal, as converted above, embedded with the data signal, also as converted
- 25 above, in such a way that the data signal is synchronized with the audio signal. This synchronization preferably occurs by embedding, into one of the frames of the audio sequence, one of the data segments.

More particularly, the encoding process occurs preferably in the following manner. First, the audio signal is mapped into a plurality of audio segments (step 105). These audio segments, which are similar in nature to the above-described data segments, preferably correspond to one beat of the song. After the control signal is encoded and included within the data signal, each audio segment is packed into one of the frames of the audio sequence (step 110). Additionally, one of the data segments is packed into the frames of the audio sequence, such that the data segment corresponds to the audio segment packed into the frame of the audio sequence.

- 10 Preferably, the sequence of encoding is such that the data segments are embedded into the audio sequence in the private bit in the header field first (step 115). Upon filling that private bit, any future data segments are preferably embedded into the private bit in the main data field (step 120). If both of the private bits are filled, then any remaining data segments would be  
15 embedded into the ancillary data field (step 125).

It should be noted that the data signal is embedded into a lower level of the audio sequence (i.e., the fields and sub-fields), as opposed to a high level, such as within the frames themselves. In this way, all the embedded data will be supported by standard MPEG decoders, and no additional  
20 circuitry will be needed to capture the data.

In operation, for example, assuming the musical composition to be the musical composition "Layla," the audio sample would contain the music to the composition. The data sample would be the lyrics to the composition. Both samples are then converted to, for example, MP3 formats. During the  
25 encoding process, the lyrics to the song would be separated in accordance with the beat or tempo of the music. Thus, the first line of the song ("What would you do if you get lonely?") would be separated into the first nine beats of the music, one for each syllable. The data signal and the audio signal would then be encoded to form the audio sequence in a manner such that the  
30 frame containing the first beat would also contain the first word, and so on.

- 10 -

- Alternatively, in an alternative embodiment, and in lieu of encoding the audio sequence with the data, the audio sequence may be encoded with a series of pointer signals. The pointer signals refer to the data signal, which, in this embodiment, is stored in a separate file. Additionally, the pointer signals
- 5 reference the data signal in accordance with the instructions contained within the control signal, and are synchronized in the same way as the data signal is synchronized in the preferred embodiment (i.e., the pointer signals would refer to the data signals in such a way that the audio sequence is synchronized with the data signal). In this case, the audio sequence would be
- 10 encoded in such a manner that the frame containing the first beat would also contain a pointer referencing the separate data file.

After the encoding process has taken place, the audio sequence may be outputted to either a karaoke player, or to any presently known storing medium for play at a future time (step 130). With reference to FIGS.1-7, a

15 method of outputting an audio signal having a synchronized data signal is provided. The audio sequence, encoded preferably in the manner set forth above, is provided (step 200). Contained within the audio sequence is a compressed audio signal. This compressed audio signal corresponds to the audio signal, described above, which contains the song portion of the musical

20 composition. Additionally provided is a compressed data signal, corresponding to the lyrical portion of the musical composition. The compressed data signal may be located within the audio signal, or within a separate data file (in which case, the audio sequence may include the pointer signals), as described above. At this point, the compressed data signal is

25 currently synchronized with the compressed audio signal. The compressed data signal is then unpacked and stored in a buffer (steps 205, 210, 215). The compressed audio signal is also unpacked. Both signals are then synchronously outputted to an output device, which may be, for example, a karaoke player system (steps 220, 225). Alternatively, the output device may

30 be a speaker, a stereo system, a video system or any other similar device.

- 10 -

- 11 -

- Turning now to a discussion of the apparatus, FIG. 8 shows a block diagram of an MP3 karaoke player device. Referring to FIG. 8, in conjunction with FIGS 1-7, the Interface Port 50 preferably interfaces to an external storage source, preferably through a docking station or cable. The Interface Port 50 is used to transfer ".mp3" files from the external source to the karaoke player device to be stored in the karaoke player device's Flash Memory 52. The external storage source may be a Personal Computer or other similar external device.
- 5

- The Flash Memory 52 is used to store one or more ".mp3" files to be played by the MP3 karaoke player. This type of memory can be overwritten with new information, but will "remember" any files that are stored in it until it is overwritten on purpose.
- 10

- The Memory Controller 54 is used to coordinate the interface between the Interface Port 50 and the Flash Memory 52, between the Flash Memory 52 and the MP3 Decoder 56, and between the Flash Memory 52 and the LCD controller 58. Additionally, the Memory Controller 54 is preferably used to interface to the person using the karaoke player device through the Button Controls 60.
- 15

- The MP3 Decoder 56 provides the function as described above. That is, decodes the MP3 karaoke file, (i.e., the ".mp3 file"), and outputs audio data to the Audio Mixer 62 and karaoke data to the LCD/karaoke Control 58.
- 20

- The LCD/karaoke Control 58 has several functions. First, it controls the LCD display to display text and lyrics, highlight words, and scroll lines of text. The LCD/Karaoke Control 58 also sends video cues received from the MP3 Decoder 56 to the Video Out Cue Jack 64 for external processing. Finally, it controls the Audio Mixer 62 to allow the person using the device's voice to over-ride the singers' voice in the original song.
- 25

- 11 -

- The Button Controls 60 allow the person using the device to control operation of the karaoke player device. Preferably, the button controls 60 include buttons for Play, Forward, Reverse, Pause, Stop, as well as other basic functions. The button controls 60 allow the user to select a specific
- 5 song to play and/or sing along with, skip songs, pause or otherwise manipulate the songs according to the user's desires.
- The Video Out Cue Jack 64 is provided to interface with an external device controlling the display of a music video. It is also used to send signals being decoded by the MP3 decoder 56 to this external device to sequence
- 10 the music video along with the file being played by the MP3 karaoke player.
- The LCD Display 66 provides the visual interface to the person using the karaoke player device. The LCD display 66 is large enough and flexible enough to display several rows of text, highlight text, scroll lines of text, etc. The LCD display 66 also provides karaoke functionality. The display 66 is
- 15 preferably flexible enough to display characters in many languages, as the song playing may be in a different language than the display shows.
- The Audio Mixer 62 is used to mix the source audio provided by the MP3 Decoder 56 with the voice of the person using the device from the microphone 68. The user's voice over-rides the singer's voice in the original
- 20 audio. The output of the Audio Mixer 62 is preferably sent to both a Headphone Jack 70 and an Audio Out Jack 72, preferably through a Digital to Analog Converter 74.
- Finally, the Microphone 68 allows the person using the device to sing along with the musical composition as it is played, guided by the lyrics
- 25 displayed on the LCD Display 66.
- It should be appreciated that the embodiments described above are to be considered in all respects only illustrative and not restrictive. The scope of the invention is indicated by the following claims rather than by the foregoing description. All changes that come within the meaning and range of
- 30 equivalents are to be embraced within their scope.

**WE CLAIM:**

1. A method of encoding an audio sequence with synchronized data, comprising the steps of:
  - 5 providing an audio sample and a data sample;
  - converting the audio sample into an audio signal;
  - converting the data sample into a data signal, the data signal including a plurality of data segments; and
  - encoding the audio signal with the data signal to form an audio sequence, the audio sequence including a plurality of frames, each frame including at least one field for receiving at least one data segment of the data signal.
- 10 2. The method of Claim 1, wherein the data signal further includes a control signal; and further comprising the step of:
  - 15 encoding the audio sequence in accordance with instructions contained within the control signal.
- 20 3. The method of Claim 2, further comprising the step of outputting the audio sequence.
4. The method of Claim 1, wherein the audio sequence is provided in a format selected from the group of formats consisting of MPEG 1/2 Layer 1/2, AC-3, WMA, AAC, EPAC, Liquid and G-2 formats.
- 25 5. The method of Claim 1, wherein the data sample further includes text data.
- 30 6. The method of Claim 1, wherein the data sample further includes video data.

- 14 -

7. The method of Claim 1, wherein the audio sample comprises a song.

5 8. The method of Claim 1, wherein the audio sample comprises spoken voice.

9. The method of Claim 1, wherein the encoding process further comprises the following steps:

- 10 mapping the audio signal into a plurality of audio segments; encoding a control signal, the control signal being included within the data signal; packing each audio segment into one of the frames of the audio sequence; and
- 15 packing each data segment into one of the frames of the audio sequence containing a corresponding audio segment in accordance with instructions contained within the control signal.

10. A program for encoding an audio sequence with synchronized data from a data signal, comprising:
- 20 computer readable program code that provides an audio sample and a data sample;
- computer readable program code that converts the audio sample into an audio signal;
- 25 computer readable program code that converts the data sample into a data signal, the data signal including a plurality of data segments; and
- computer readable program code that encodes the audio signal with the data signal into an audio sequence, the audio sequence including a plurality of frames, each frame including at least one field for receiving at least
- 30 one data segment of the data signal.

- 14 -

- 15 -

11. A method of encoding an audio sequence with synchronized data, comprising the steps of:
    - providing an audio sample and a data sample;
    - 5 converting the audio sample into an audio signal;
    - converting the data sample into a data signal, the data signal including a plurality of data segments; and
    - encoding the audio signal with a plurality of pointer signals to form an audio sequence, each pointer signal referencing at least one data
  - 10 segment of the data signal.
- 
12. The method of Claim 11, wherein the data signal further includes a control signal; and further comprising the step of:
    - encoding the audio sequence in accordance with instructions contained within the control signal.
- 
13. The method of Claim 12, further comprising the step of outputting the audio sequence.
- 
- 20 14. The method of Claim 11, wherein the audio sequence is provided in a format selected from the group of formats consisting of MPEG 1/2 Layer 1/2, AC-3, WMA, AAC, EPAC, Liquid and G-2 formats.
  15. The method of Claim 11, wherein the data sample further includes text data.
  - 25 16. The method of Claim 11, wherein the data sample further includes video data.

- 15 -

- 16 -

17. The method of Claim 11, wherein the audio sample comprises a song.

5 18. The method of Claim 11, wherein the audio sample comprises spoken voice.

19. The method of Claim 12, wherein the encoding process further comprises the following steps:

10 mapping the audio signal into a plurality of audio segments;  
encoding a control signal, the control signal being included  
within the data signal;  
packing each audio segment into one of the frames of the audio  
sequence; and  
15 packing into each audio segment one of the pointer signals,  
each pointer signal referencing one of the data segments of the data signal.

20. A program for encoding an audio sequence with synchronized  
data, comprising:

20 computer readable program code that provides an audio sample  
and a data sample;  
computer readable program code that converts the audio  
sample into an audio signal;  
computer readable program code that converts the data sample  
25 into a data signal, the data signal including a plurality of data segments; and  
computer readable program code that encodes the audio signal  
with a plurality of pointer signals to form an audio sequence, each pointer  
signal referencing at least one data segment of the data signal.

- 16 -

21. A method of outputting an audio signal having a synchronized data signal, comprising the steps of:
- providing an audio sequence with synchronized data, the audio sequence including a compressed audio signal;
- providing a compressed data signal, the compressed data signal being synchronized with the compressed audio signal;
- unpacking the compressed data signal;
- storing the data signal in a buffer;
- 10            unpacking the compressed audio signal from the audio sequence; and
- outputting the audio signal and the data signal to an output device.
- 15            22. The method of Claim 21, further comprising the step of unpacking the compressed data signal from the audio sequence.
- 20            23. The method of Claim 21, wherein the audio sequence further includes a plurality of pointer signals, each pointer signal referencing the compressed data signal.
24. The method of Claim 21, wherein the audio sequence is in MP3 format.
25. The method of Claim 21, wherein the audio signal is a signal selected from the group consisting of a song and a spoken voice, and wherein the data signal is a signal selected from the group consisting of text and a spoken voice.

26. The method of Claim 21, wherein the output device is a device selected from the group consisting of a speaker, a stereo system, a karaoke system and a video system.

5

27. A program for outputting an audio signal having a synchronized data signal, comprising:

computer readable program code that provides an audio sequence with synchronized data, the audio sequence including a

10 compressed audio signal;

computer readable program code that provides a compressed data signal, the compressed data signal being synchronized with the compressed audio signal;

15 computer readable program code that unpacks the compressed data signal;

computer readable program code that stores the data signal in a buffer;

computer readable program code that unpacks the compressed audio signal from the audio sequence; and

20 computer readable program code that outputs the audio signal and the data signal to an output device.

28. The program of Claim 27, further comprising:

computer readable program code that unpacks the compressed data signal from the audio sequence.

29. The method of Claim 27, wherein the audio sequence further includes a plurality of pointer signals, each pointer signal referencing the compressed data signal.

30

1/5

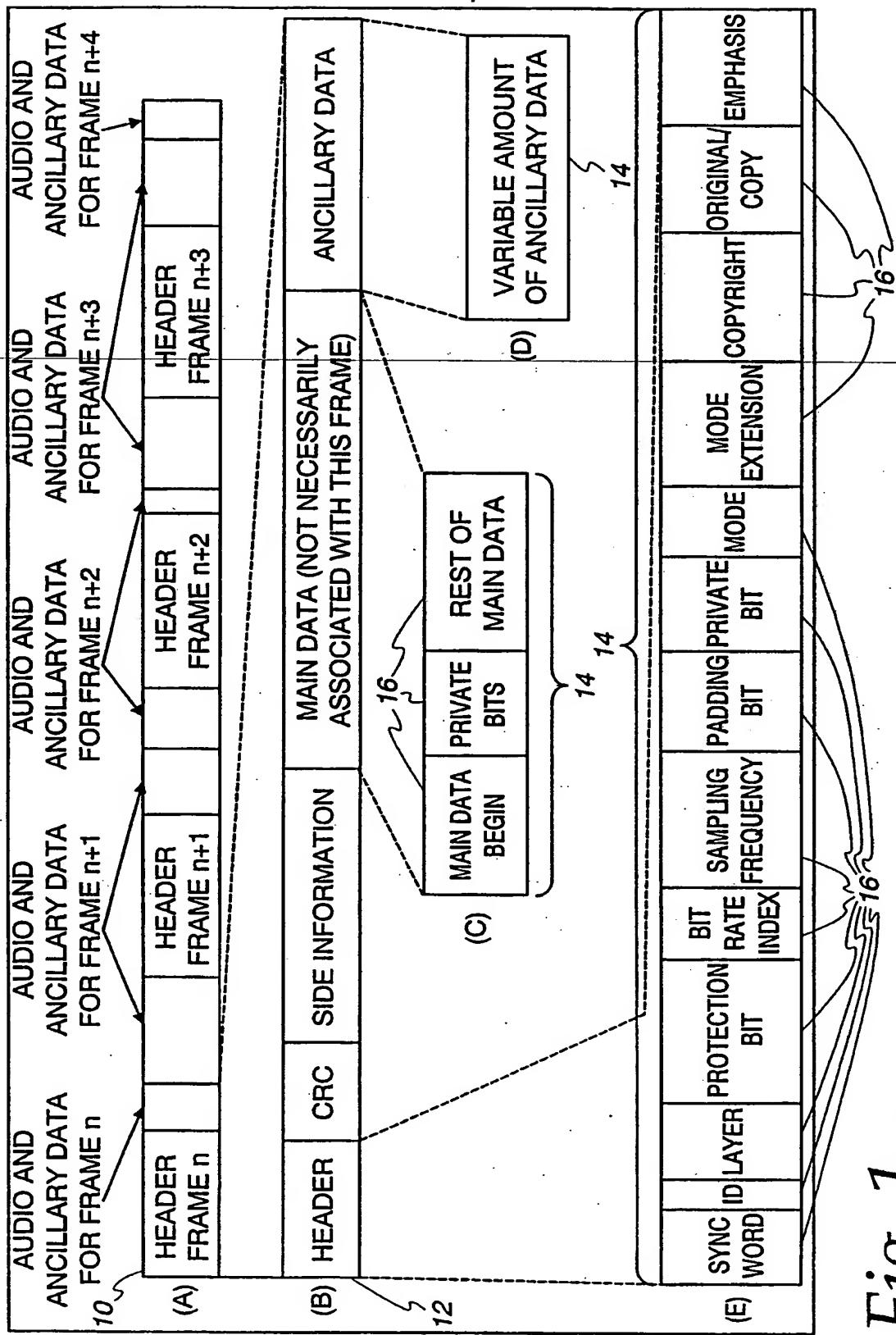
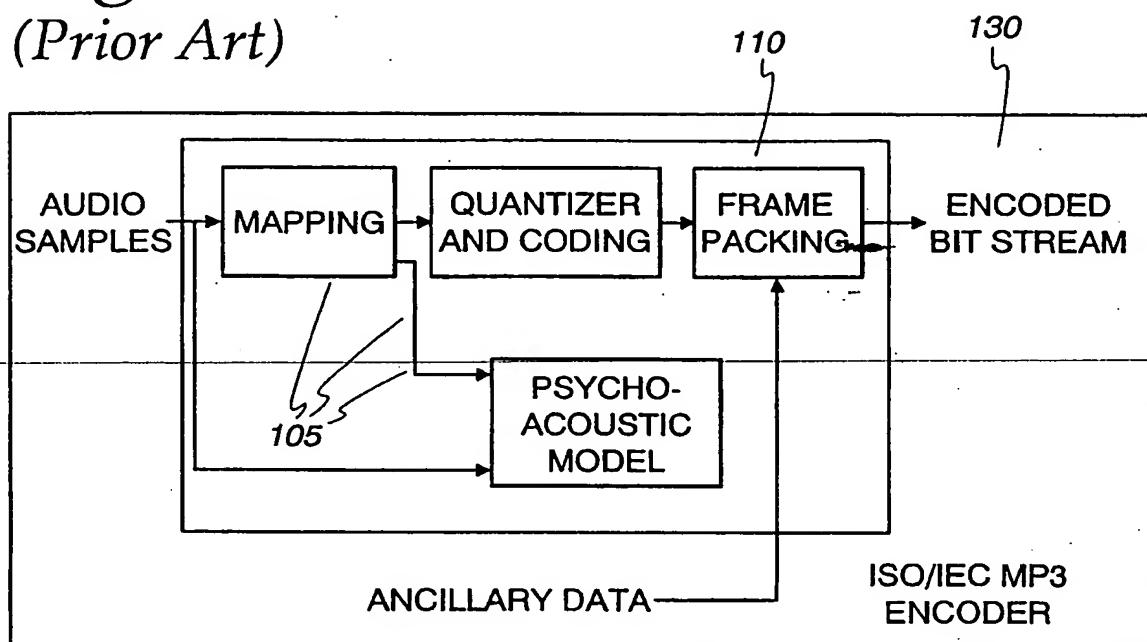


Fig. 1 (Prior Art)

*Fig. 2*  
(Prior Art)

2/5



*Fig. 3*

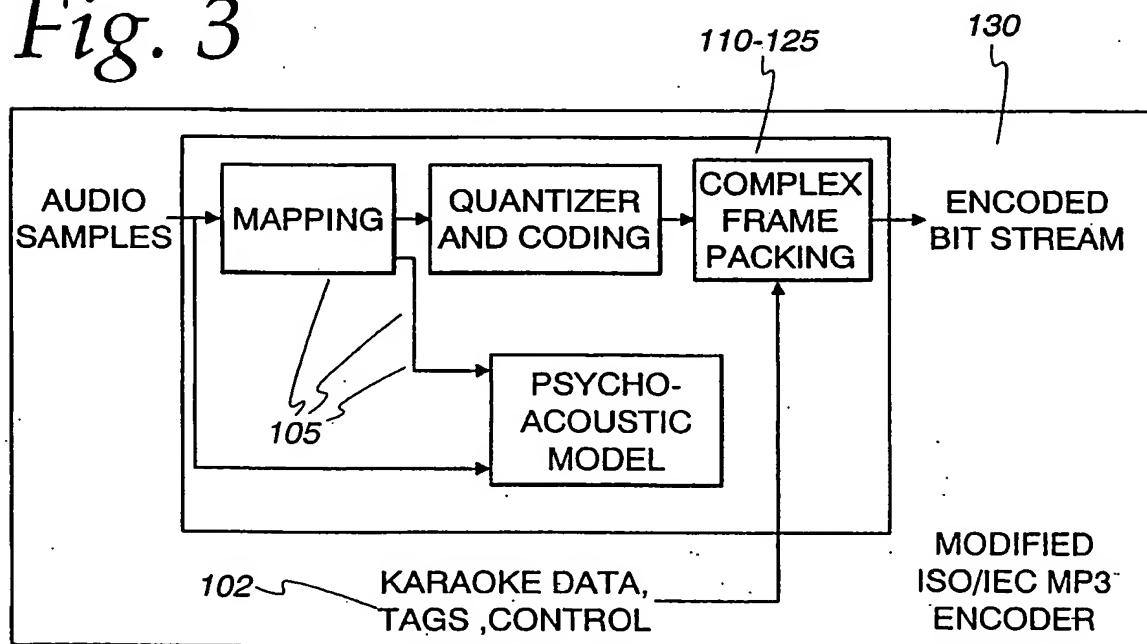


Fig. 4

3/5

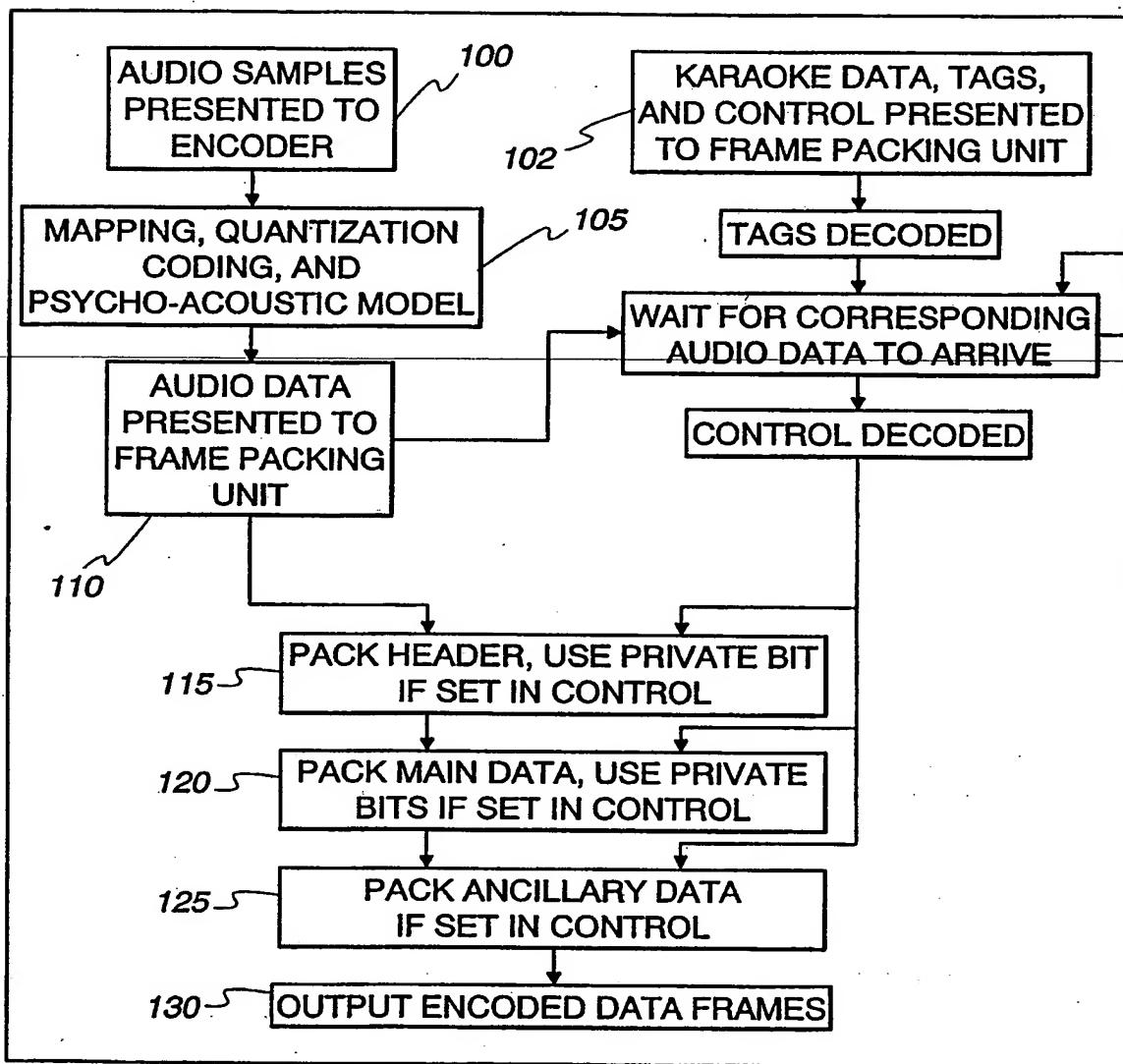


Fig. 5 (Prior Art)

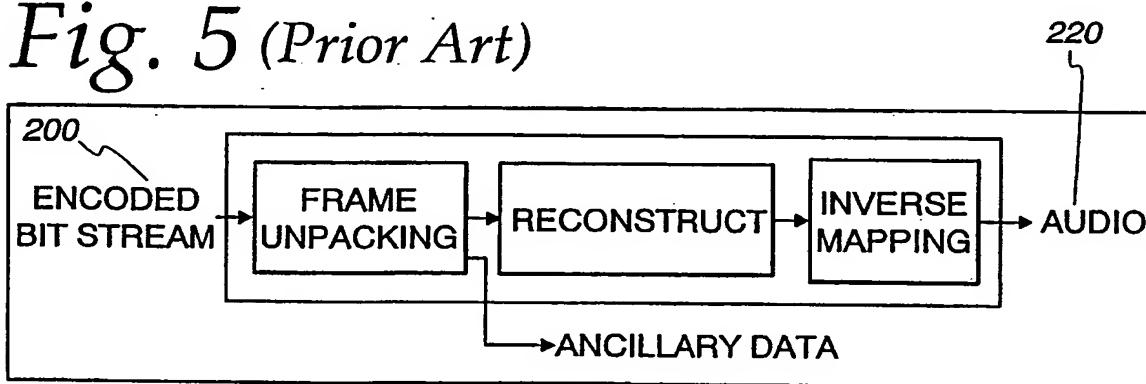


Fig. 6

4/5

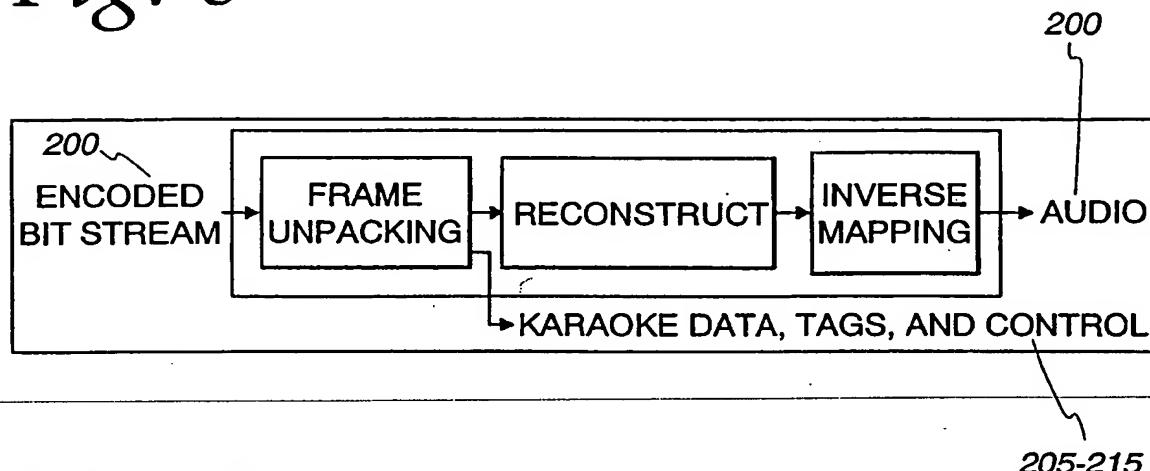
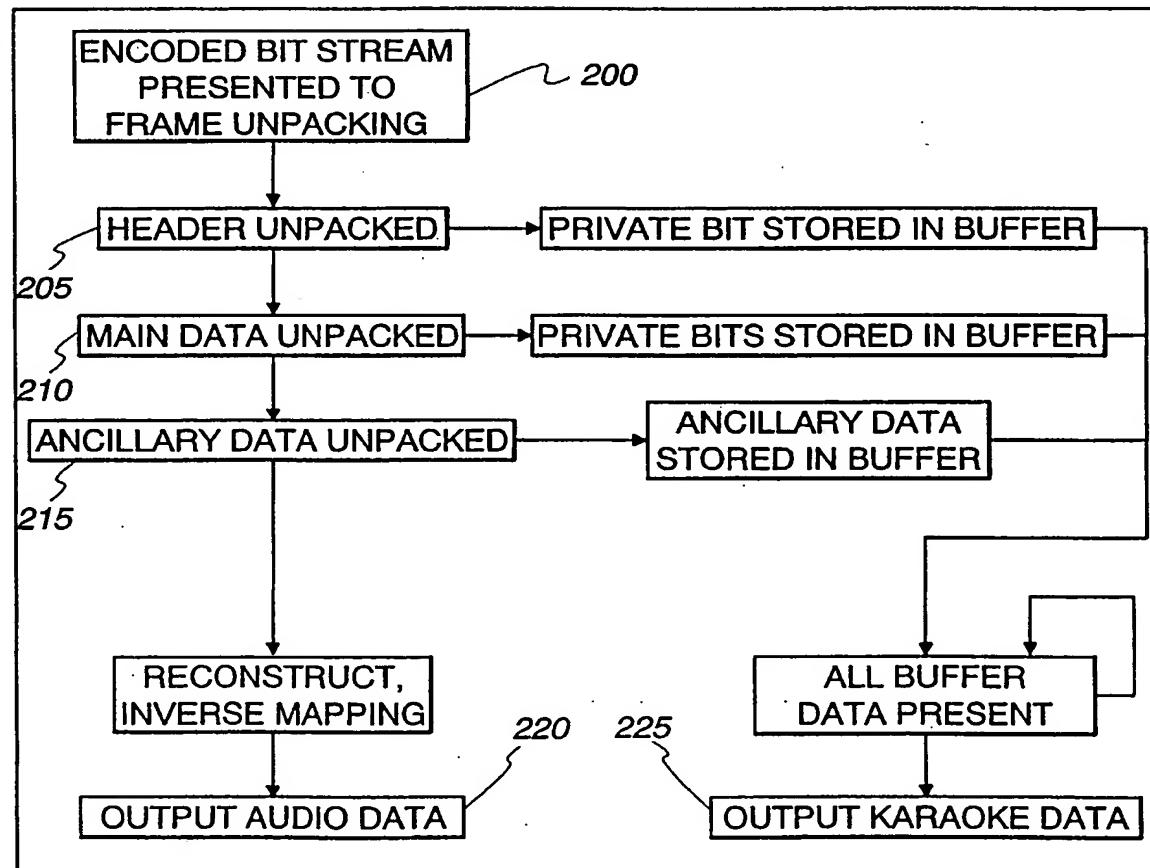


Fig. 7



5/5

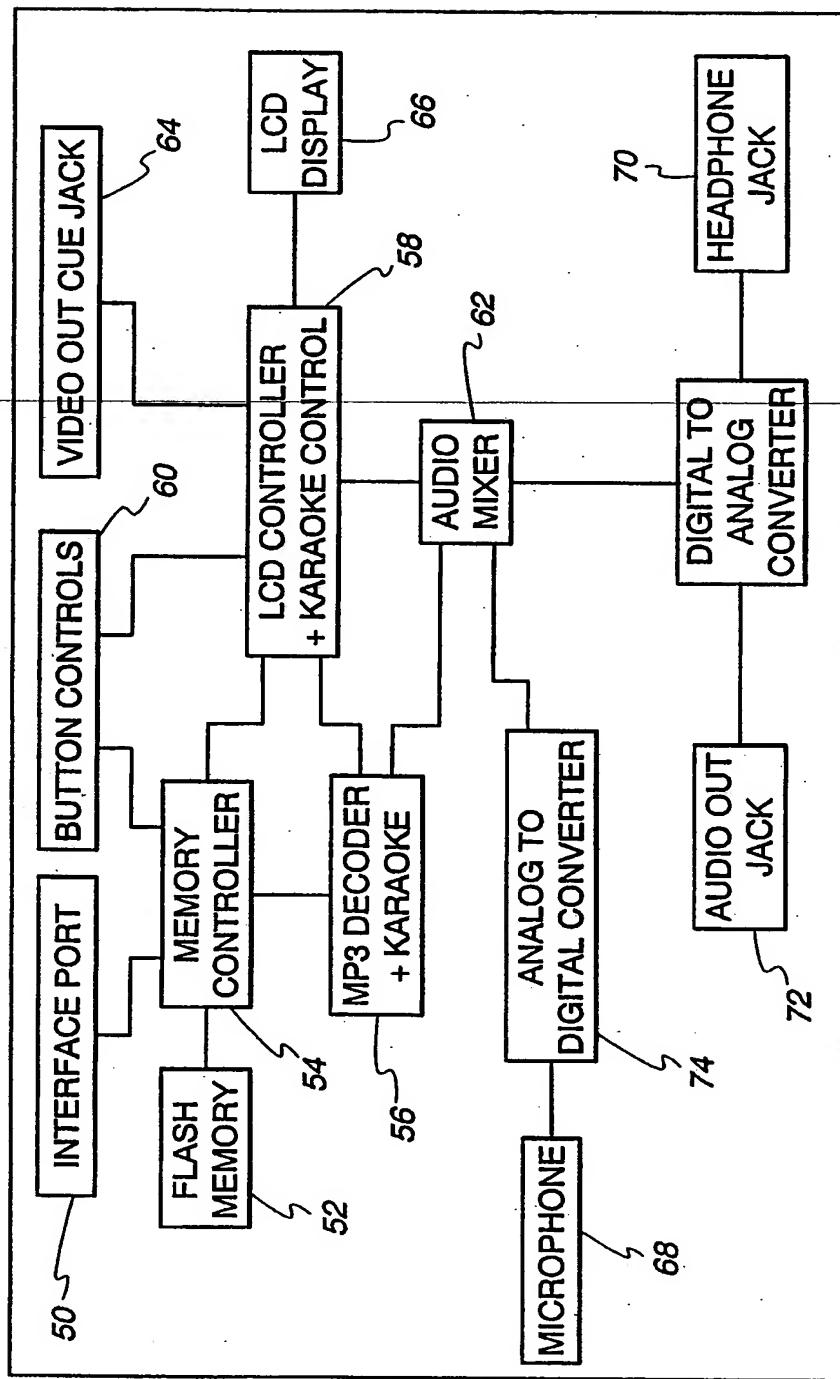


Fig. 8

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US00/31338

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : G10L 19/00; H04J 1/02  
US CL : 704/201, 236, 270.1, 500; 370/493; 709/236

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
U.S. : 704/201, 270, 270.1, 500; 370/493; 709/236

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
BRS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No.                          |
|------------|--|--|
| X          | GB 2 323 760 A (NEC CORPORATION) 30 September 1998 (30.09.98), entire document.  | 1, 3, 5, 8, 9, 10, 21, 22, 27, 28              |
| Y          | US 5,281,985 B1 A (CHAN) 25 January 1994 (25.01.94), column 1, column 4, column 8.   | 26   |
| X          | US 5,732,216 B1 B1 (LOGAN et al.) 24 March 1998 (24.03.98), columns 1-5, columns 14-15, column 19, column 28.  | 21, 22, 23, 25, 27, 28, 29                     |
| Y          |  | 7, 8, 17, 18                                   |
| Y          | US 5,777,997 A (KAHN et al.) 07 July 1998 (07.07.98), entire document.   | 7, 17  |
| X          | US 5, 856,973 B1 (THOMPSON) 05 January 1999 (05.01.99), entire document  | 1, 2, 3, 5, 6, 9-13, 15, 16, 19-24, 27, 28, 29 |
| ---        |  | 4, 7, 8, 14, 17, 18, 26                        |
| Y          | CRAVOTTA, NICHOLAS. The Internet-audio Revolution. EDN. 03 February 2000 (03.02.00), vol.45, iss.3, pp. 101-107, especially pages 101-102, page 105. | 4, 14  |

|                          |   |                          |  |
|--------------------------|---|--------------------------|--|
| <input type="checkbox"/> | Further documents are listed in the continuation of Box C.  | <input type="checkbox"/> | See patent family annex.   |
| *                        | Special categories of cited documents:  | "T"                      | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| "A"                      | document defining the general state of the art which is not considered to be of particular relevance  | "X"                      | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| "E"                      | earlier application or patent published on or after the international filing date   | "Y"                      | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L"                      | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&"                      | document member of the same patent family  |
| "O"                      | document referring to an oral disclosure, use, exhibition or other means  |                          |  |
| "P"                      | document published prior to the international filing date but later than the priority date claimed  |                          |  |

|  |   |
|--|---|
| Date of the actual completion of the international search<br>27 February 2001 (27.02.2001)   | Date of mailing of the international search report<br>17 APR 2001       |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703)305-3230 | Authorized Officer<br>William R. Korzuch<br>Telephone No. (703)305-4700 |

---

**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**